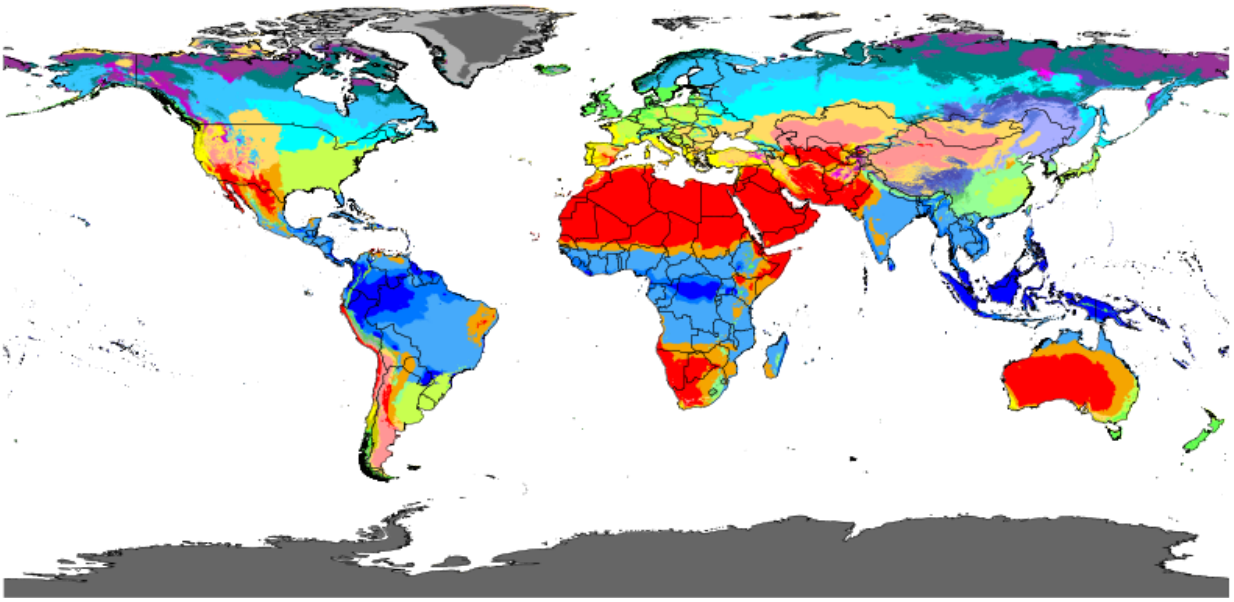


Explore Weather Trends

07.16.2020

—

Amber Cross
Data Analyst Nanodegree Program

Overview

In this first project, I used the provided weather database to extract data for specific cities, then identify and analyze the apparent trends.

Goals

To demonstrate:

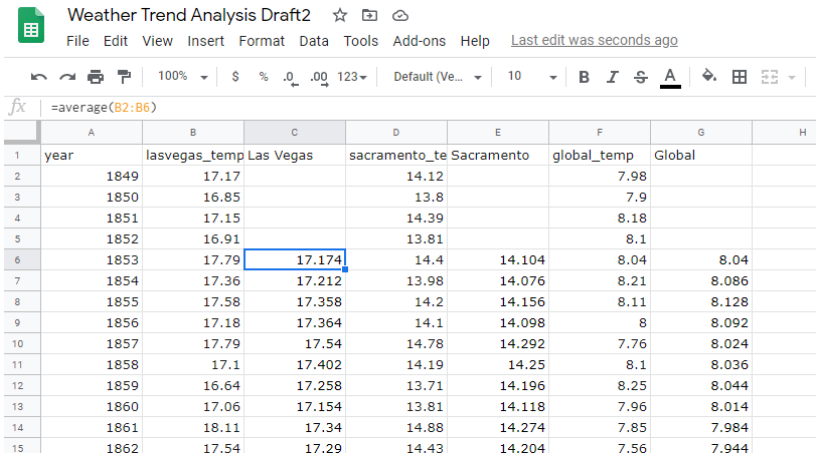
1. Familiarity with running SQL queries to extract relevant data.
2. Ability to manipulate data to calculate a moving average in order to populate a line chart with easily interpreted trends.
3. Observations and analysis of the trends.

Methodology

I started by running the following query in the work space for both my current city, Las Vegas, my hometown of Sacramento, and the global temperature data.

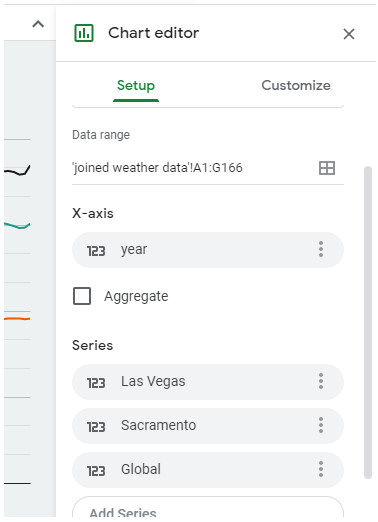
```
SELECT
        city_data1.year,
        city_data1.avg_temp AS lasvegas_temp,
        city_data2.avg_temp AS sacramento_temp,
        global_data.avg_temp AS global_temp
FROM city_data AS city_data1
JOIN city_data AS city_data2
ON city_data1.year = city_data2.year
JOIN global_data
ON city_data1.year = global_data.year
WHERE city_data1.city LIKE 'Las Vegas' AND city_data2.city LIKE 'Sacramento'
ORDER BY city_data1.year;
```

I opened the downloaded query results within Google Sheets to further manipulate the data. Taking a ten-year average of the data overly smoothed the line chart, so I settled on a five-year moving average, by using the AVERAGE function over the first five years/cells of the temperature column and then copying down through the dataset, for each element: Las Vegas, Sacramento, and global. I created the line chart with the results and labeled it appropriately, displayed in my findings.

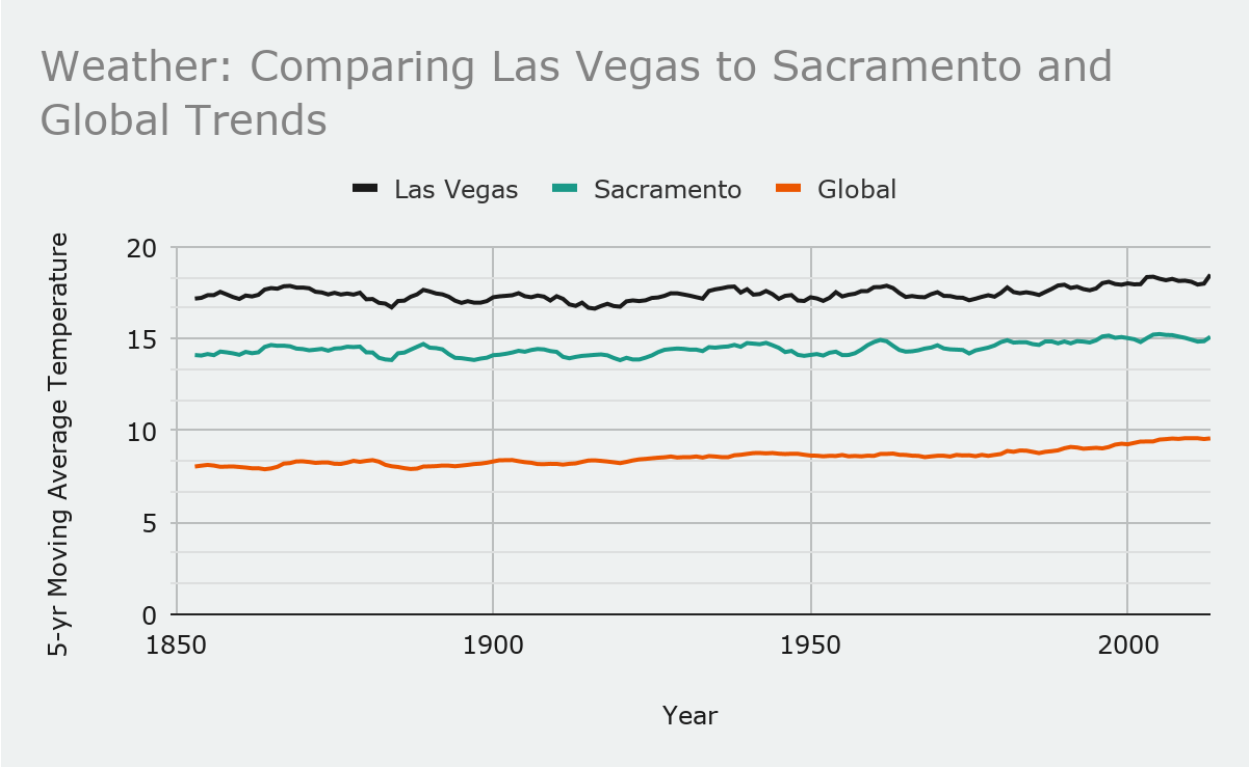


	A	B	C	D	E	F	G	H
1	year	lasvegas_temp	Las Vegas	sacramento_te	Sacramento	global_temp	Global	
2	1849	17.17		14.12		7.98		
3	1850	16.85		13.8		7.9		
4	1851	17.15		14.39		8.18		
5	1852	16.91		13.81		8.1		
6	1853	17.79	17.174	14.4	14.104	8.04	8.04	
7	1854	17.36	17.212	13.98	14.076	8.21	8.086	
8	1855	17.58	17.358	14.2	14.156	8.11	8.128	
9	1856	17.18	17.364	14.1	14.098	8	8.092	
10	1857	17.79	17.54	14.78	14.292	7.76	8.024	
11	1858	17.1	17.402	14.19	14.25	8.1	8.036	
12	1859	16.64	17.258	13.71	14.196	8.25	8.044	
13	1860	17.06	17.154	13.81	14.118	7.96	8.014	
14	1861	18.11	17.34	14.88	14.274	7.85	7.984	
15	1862	17.54	17.29	14.43	14.204	7.56	7.944	

ABOVE: calculating five-year moving average BELOW: creating line chart



Findings


Weather: Comparing Las Vegas to Sacramento and Global Trends
Las Vegas
Sacramento
Global
5-yr Moving Average Temperature
20
15
10
5
0
1850
1900
1950
2000
Year

Conclusions

I. Las Vegas is warmer than both Sacramento and the global average

This is not surprising, since Las Vegas is in the desert and known for its sweltering summertime. I was a little more surprised to observe that Sacramento is much closer to Las Vegas than the global average. It is also interesting to observe how closely Sacramento and Las Vegas mirror each other: the correlation coefficient between Sacramento and Las Vegas is .89, versus .74 with Sacramento and the global data (calculated by using the CORREL function over the five-year moving average data ranges for each element).

II. Globally, there is a warming trend

Reviewing the data for the moving average for global data shows a coefficient of .91, which is a strong correlation. There is an average increase of .11% for the five-year moving average over the global dataset.

III. Is Las Vegas warming?

When I originally pulled data for Las Vegas, there were gaps in the data and a few really hot years early in the dataset. I tried to smooth over the gaps by taking the average temperature to that point in their place using COALESCE. It was difficult to see any trend; the average temperature for Las Vegas really seemed to hover more than move in any direction until I honed in on only the last 100 years of the dataset. By viewing the data again in the window created by a query that joined Las Vegas with the global data and Sacramento, there is a clearer warming trend. What I thought was a logical solution to early gaps seemed to actually obscure the trends.

The correlation coefficient comparing years to the moving average temperature for Las Vegas is only .49 and the average increase only .05%. Perhaps this would make sense given that deserts are already relatively hot, however Sacramento has only an average increase of .046%. Some of the largest five-year-average increases globally were accompanied by decreases in Las Vegas and Sacramento (see 1998, 2005, 1973).

	A	B	C	D	E	F	G	H	I	J
	year	lasvegas_temp	Las Vegas	LV Increase	sacramento_te	Sacramento	SAC increase	global_temp	Global	Global Increase
	1973	17.03	17.232	-0.4621072089	14.58	14.398	-0.1110031913	8.95	8.67	1.001863933
	2005	17.95	18.252	-0.6098889131	15.17	15.244	0.1971868016	9.7	9.506	1.06315118
	1869	17.63	17.774	-0.5038065383	14.57	14.454	-0.8369923161	8.43	8.318	1.09382596
	1866	17.95	17.712	-0.180342651	14.67	14.606	-0.2867285636	8.29	8.024	1.108870968
	1878	17.17	17.392	-0.320953691	14.37	14.538	-0.1236603462	8.83	8.348	1.163354338
	1961	17.48	17.8	0.0337192312	14.65	14.926	0.7696462328	8.8	8.722	1.206776514
	1938	17.23	17.828	0.1123090746	14.35	14.654	0.5765271105	8.86	8.652	1.216658868
	1990	17.64	17.928	0.23482053	14.73	14.842	0.6783340117	9.23	9.034	1.278026906
	1889	17.8	17.648	1.413630617	14.81	14.704	1.044529962	8.32	8.038	1.387487386
	1998	16.71	17.966	-0.6634966272	14.38	15.042	-0.7783641161	9.52	9.23	1.428571429
	1981	18.63	17.78	1.681345076	15.22	14.908	0.6753106429	9.17	8.884	1.880733945
	1867	18.2	17.844	0.7452574526	14.46	14.612	0.04107900863	8.44	8.2	2.193419741
	1849	17.17			14.12			7.98		

ABOVE: capture of data sorted by global increase

IV. Meaningful data

Trying to make sense of the contradictions above took me to climate.nasa.gov which displayed more seasonal data; fluctuations specifically during hottest days and coldest nights of those seasons. The warming trends are much more obvious when viewed this way, with troubling increases in the arctic nights during cool seasons and mid-latitude region days during warm seasons. I conclude that annual temperature averages are not the most effective way to measure and study warming trends, though it is still apparent.